

AD737836

NRL MEMORANDUM REPORT NO. 458

SHIFT IN PROBLEM SOLVING

Scarvia B. Anderson

15 April 1955

Systems Analysis Branch
Applications Research Division
Naval Research Laboratory
Washington 25, D. C.

ACKNOWLEDGEMENTS

Dr. Omar K. Moore assisted materially in the initial stages of the experiment reported here. Dr. Moore, now at Yale University, was formerly head of the Problem-Solving project, Systems Analysis Branch. The author is also indebted to Dr. A. H. Ryan, Dr. E. P. Buckley, Mr. L. M. Chauvette, Mr. F. J. McGrane, Dr. H. W. Sinaiko and Mr. D. F. Wilson of the Systems Analysis Branch, LT A. E. Hickey, ONR, and Dr. T. G. Andrews, University of Maryland, for helpful criticisms during the preparation of this report; to Mrs. Doris Bridges for clerical assistance; and to the Center Commander, Bainbridge NTS, for providing quarters, subjects, and services during experimental runs there.

. . .

A condensed version of this report was presented at the meeting of the Eastern Psychological Association, Philadelphia, April 1955.

CONTENTS

| | |
|--------------------------|-----|
| ABSTRACT | ii |
| PROBLEM STATUS | iii |
| AUTHORIZATION | iii |
| INTRODUCTION | 1 |
| METHOD | 2 |
| RESULTS | 8 |
| DISCUSSION | 15 |
| REFERENCES | 19 |

ABSTRACT

The performance of human subjects forced to shift back and forth every 3 min. from one problem-solving task to another was compared with the performance of subjects allowed to work as long as 30 min. on one problem before going on to a second problem. Some interference effects were predicted for shift subjects.

Using a task drawn from the calculus of propositions, it was possible to specify not only effects on over-all problem-solving performance but also effects on certain intermediate behaviors and generalized activity level. The results may be summarized as follows:

1. Shift and non-shift subjects did not differ significantly with respect to number of solutions obtained, time for solutions obtained, number of errors, and activity level.
2. Shift subjects engaged in more redundant activity than non-shift subjects.
3. Shift subjects were more variable than non-shift subjects in terms of times for solutions, errors, and redundant activity.
4. Some of the attitudes of shift subjects toward the experimental procedure were less favorable than those of non-shift subjects.

PROBLEM STATUS

The experiment reported here is one of a continuing series in which aspects of human problem-solving behavior and variables which affect it are being systematically explored.

AUTHORIZATION

NRL Problem Y03-01

Project NR 593-010

INTRODUCTION

Forced shift from one problem-solving task to another has been shown to result in negative or positive transfer, depending upon kind and similarity of tasks, amount of practice or training, intelligence, and number of shifts (1, 2, 4, 5, 7, 12).¹ Positive transfer has been explained in terms of the development of a "higher-order pattern of response" (8) or a "more comprehensive mental set" (7), such that the separate responses demanded by the two tasks are essentially combined into one integrated pattern of responses. Harlow (6) presents evidence for such higher-order habits in problem solving by apes, children, and adults. Negative transfer, on the other hand, is generally attributed to the interference of responses in one task on responses in the other task. Youtz (12) discusses negative transfer in terms of the incompatibility of problem-solving hypotheses in the two tasks.

In the present experiment, human subjects were forced to shift back and forth from one complex problem to another complex problem, similar with respect to the symbols used and the problem-solving rules which could be employed. From analogy with experiments on retroactive inhibition and from a knowledge of the results of Jersild (7) who used both similar and dissimilar tasks in shift experiments, some interference effects could be expected. We were concerned with specifying the nature of these effects.

Would the effects of working first on one problem for a short period of time, then on another problem, then returning to the first problem, etc. show up in terms of

¹Numbers in parentheses refer to references listed at the end of the report.

over-all measures of performance (i.e., number of problems solved and time required for solutions)? Or would the effects be more subtle, reflected perhaps in instances of redundant activity and errors which characterize search behavior in problem solving? Would activity level (number of problem steps attempted per unit time) increase or decrease? And what would be the attitudes of subjects toward working in a shift situation? Our operating hypotheses, stemming from the prediction of interference, were that subjects forced to shift back and forth from one problem to another would solve fewer problems, take longer times for the problems they did solve, engage in more redundant activity, and make more errors than subjects who were not forced to shift back and forth from one problem to another.

METHOD

Subjects. The subjects were 32 navy enlisted men, rather homogeneous with respect to age (\bar{X} = 20.1 years, σ = 2.9), number of years of school completed (\bar{X} = 11.6 years, σ = 1.5), and General Classification Test scores (\bar{X} = 60.3 or approximately one standard deviation above the population mean of 50; σ = 3.4). No subjects had had any previous experience with the kind of problem-solving task used in the experiment. For the experimental test of the effects of shift on problem-solving behavior, 16 subjects were assigned at random to the shift condition, 16 to the non-shift condition.

Material. The calculus of propositions, one of the calculi of symbolic logic, served as the source of problems used in the experiment. Each problem consists of a set of premises and a conclusion to be deduced from

those premises. The steps which a subject can take in attempting to reach the conclusion are governed by a set of 12 transformation rules. An example problem is presented in Table 1 (p. 4). In Sector I are the premises, and in Sector II is the conclusion to be deduced from those premises. In Sector III are the steps taken by a hypothetical subject in an attempt to deduce the conclusion. In Sector IV are the "justifications" for each step taken; e.g., Line 5 was obtained by applying Rule 7 to Line 4. Problems were presented to subjects as problems in "cracking codes." No meanings were ascribed by the experimenter to the letters ('P,' 'Q,' 'R,' 'S,' and 'T'), connectives ('v,' '∧,' '∨,' and '~'), and punctuation (parentheses and brackets) used in problems; the rules were identified to subjects only by number. More detailed accounts of the adaptation of the calculus as a source of tasks and the experimental advantages the tasks offer are presented by Moore and Anderson (10).

Ten problems similar to the one shown in Table 1 were used as practice problems for the present experiment. Eight additional problems were used to test the effects of shift on problem-solving behavior. The latter problems were divided into four pairs, the two problems in each pair being approximately matched with respect to complexity (e.g., number of premises and minimum number of steps required for solution).

Instructions. Instructions on the use of the 12 rules of the "code" were presented to subjects by means of long-playing records. Each subject was provided with a rule sheet summarizing the 12 rules; a deck of 12 plastic cards, each card containing a rule and several

Table 1
Example Problem

| Sector I | Sector II |
|---|------------------|
| <u>1.</u> $(R \cdot S) \vee P$ | |
| <u>2.</u> $(P \vee S) \supset (Q \vee R)$ | R |
| <u>3.</u> $\sim Q$ | |
| Sector III | Sector IV |
| <u>4.</u> $P \vee (R \cdot S)$ | <u>1.</u> 1 |
| <u>5.</u> $(P \vee R) \cdot (P \vee S)$ | <u>4.</u> 7 |
| <u>6.</u> $P \vee S$ | <u>5.</u> 8 |
| <u>7.</u> $\sim P \supset S$ | <u>6.</u> 6 |
| <u>8.</u> $P \vee S$ | <u>7.</u> 6 |
| <u>9.</u> $Q \vee R$ | <u>2, 8.</u> 11 |
| <u>10.</u> $\sim Q \supset R$ | <u>9.</u> 6 |
| <u>11.</u> $Q \vee R$ | <u>2, 8.</u> 11 |
| <u>12.</u> R | <u>10, 3.</u> 11 |

example expressions to which the rule could be applied; and a board equipped with magnetic symbols (letters, connectives, and punctuation) which were manipulated in accordance with instructions from the recording. The subjects were trained in groups of eight. Training periods lasted 80 min. The development of the training method and aids is discussed by Moore and Anderson (10).

Practice. Prior to the major experimental runs, each subject participated in 10 practice sessions (one problem per session) on 10 consecutive working days. Sessions lasted a maximum of 30 min. Practice sessions were run according to the same basic procedure as experimental non-shift runs except that subjects worked in groups of three on problems.

Apparatus and procedure. During experimental sessions, an individual subject sat in a semi-darkened room facing a screen. At the beginning of a session the premises and conclusion (Sectors I and II) of a problem were projected on the screen from an overhead projector. The subject's task was to call out steps which he wanted to take in an attempt to deduce the conclusion. His responses were phrased as follows: "Apply Rule 1 to Line 1" (the response which would yield Line 4 of the example problem in Table 1). The subject was provided with a sheet containing the 12 rules, so that his task was not primarily dependent on memory. The experimenter, sitting at the overhead projector and separated from the subject by a screen, reacted to the subject's responses in one of two ways: (a) If the suggestion called out by the subject was correct (i.e., if the rule specified was applicable to the line specified), the experimenter carried it out, writing in Sector III the new step generated and in Sector IV the relevant

rule-line notation. He wrote directly on the face of the projector, and the results were projected to the screen as he wrote them. (b) If the suggestion called out by the subject constituted an error (i.e., if the rule specified was not applicable to the line specified), the experimenter sounded an error buzzer, and, of course, no new step was generated.

All subjects worked individually in four experimental sessions, distributed through four consecutive working days and lasting a maximum of one hour each. Two problems were presented during each session. Order of presentation of problems was varied randomly so that eight subjects in each group received a pair of problems in a-b order and eight received them in b-a order. This arrangement was used in order to balance out possible effects on performance on the second problem of the particular problem presented first to non-shift subjects, for shift subjects would be exposed to both problems early in the session.

Each of the 16 non-shift subjects was allowed to work on one problem until he had solved it or 30 min. had elapsed (whichever occurred first); he was then presented with the second problem for the session and allowed to work on it until he had solved it or 30 min. had elapsed.

The 16 shift subjects worked 3 min. on one problem, then 3 min. on the second problem, then 3 min. on the first problem, etc. until they had solved a problem or 60 min. had elapsed. If a subject solved one problem, he was allowed to work uninterrupted on the other problem until he had solved it or worked a total of 30 min. on

it. The 3-min. intervals were timed by an automatic timer which provided an auditory signal to the experimenter to shift from one problem to another. Two overhead projectors were used, and the shift was made by turning off the projector displaying one problem and simultaneously turning on the projector displaying the other problem. (There was a brief moment when the two problems appeared to be superimposed.)

Problem-solving response measurement. The experimenter obtained from each problem-solving session a complete record of both correct steps and errors made by a subject and the time it took him to solve a problem if a solution was reached. Each correct step was evaluated as to whether it did or did not constitute a redundancy; i.e., was the same as one of the premises or as a step already taken. Two kinds of redundancy were identified: cycles and repetitions. Cycling generally involves using an expression a to get another expression b which is in turn used to get expression a again. In practice, the number of mediating steps between a and its recurrence may vary. In Table 1 (p. 4), Line 8 represents a cycle. Line 7 was obtained from Line 6; Line 7 was then used to obtain Line 8 which is identical with Line 6. Repetition is a less subtle form of redundancy. When a subject repeats, he simply calls out the same step suggestion twice. In Table 1, Line 11 represents a repetition. The hypothetical subject obtained it by exactly the same means he used to obtain Line 9; i.e., by calling out the same rule-line combination. Since no correct steps that a subject took were deleted from his problem work, and since he could go back and work in any order with any steps he had taken, redundant steps were not necessary for the problem-solving process. A more complete account of cycles and repetitions

and other kinds of response measures obtainable with the calculus task is available elsewhere (11).

Attitude measurement. Following the fourth and last experimental session, each of the 32 subjects was given a brief questionnaire designed to elicit the expression of attitudes toward the problem-solving situation. The questionnaire was patterned after a previously developed Task Experience Inventory² and consists of 15 three-option items of the following type:

As compared with other situations in which my performance was being measured, I felt that this situation was

_____ (a) definitely pleasant.

_____ (b) neither especially pleasant nor especially unpleasant.

_____ (c) definitely unpleasant.

In determining a subject's over-all score on the questionnaire, three points were assigned to each "favorable" response (e.g., choice a above), two points to each middle or neutral response, and one point to each "unfavorable" response.

RESULTS

Fig. 1 (p. 9) shows the number of problems solved under the shift and non-shift conditions during the four problem-solving sessions. The shape of the curves is probably more attributable to differential difficulty of

²The Task Experience Inventory was developed in connection with research on behavior decrement conducted at the University of Maryland Laboratory on a research contract for the Army Medical Department. It contains 19 four-option items designed to measure the amount of stress an individual feels as a result of being subjected to certain experimental conditions.

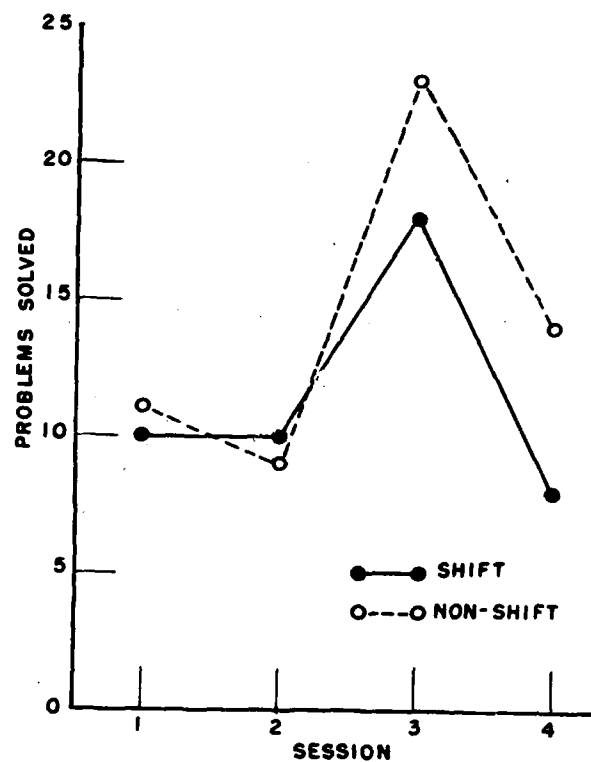


Figure 1 - Total number of problems solved by shift and non-shift subjects during each of four problem-solving sessions

the four pairs of problems than to learning. Non-shift subjects solved a total of 57 problems; shift subjects, 46. The mean number of problems solved by non-shift subjects is 3.56; by shift subjects, 2.88. This difference is not significant with the one-tail test appropriate to our initial hypothesis.

The means, t ratios, variance estimates, and F ratios for number of problems solved are shown in Table 2 (p. 11), along with the summary of results for errors, cycles, repetitions, and activity level. For these five distributions of responses, F tests indicate that in three cases the variances for shift and non-shift subjects are significantly different. In terms of errors, cycles, and repetitions, shift subjects appear to be more variable than non-shift subjects. In comparing the means for these three sets of distributions, therefore, it was necessary to modify the conventional t test, reducing df to 15 (9, p.75).

Shift subjects averaged 22.69 errors and non-shift subjects averaged 10.62 for the four problem sessions. The obtained t does not prove to be significant. Here we should note that one subject in the shift group made 94 errors during the four sessions, accounting for the great part of the difference between the two groups.

The mean number of cycles for shift subjects is 13.38; for non-shift subjects, 6.62. The t is significant beyond the .05 level. Similar results were obtained with the other kind of redundant activity: repetitions. Shift subjects averaged 2.62 repetitions; non-shift subjects, 1.25. The difference is significant beyond the .02 level. The numbers of cycles and repetitions for the two groups for the four sessions are shown graphically in Fig. 2 (p. 12).

Shift subjects made a total of 1734 (1371 correct and

Table 2
Comparison of Means and Variance Estimates

| Measure | Means | | | | Variance Estimates | | | |
|----------------|-------|-----------|-----------|------------------|--------------------|--------|-----------|-------------------------|
| | Shift | Non-Shift | \bar{t} | \underline{df} | \underline{p} | Shift | Non-Shift | $\underline{F}_{15,15}$ |
| Solutions | 2.88 | 3.56 | 1.68 | 30 | | 5.45 | 6.80 | 1.25 |
| Errors | 22.69 | 10.62 | 1.51 | 15 | | 597.03 | 82.92 | 7.20 |
| Cycles | 13.38 | 6.62 | 2.04 | 15 | < .05 | 117.45 | 23.58 | 3.97 |
| Repetitions | 2.62 | 1.25 | 2.48 | 15 | < .02 | 3.72 | 1.27 | 2.93 |
| Activity Level | 2.68 | 2.77 | .35 | 30 | | 1.55 | 2.07 | 1.33 |

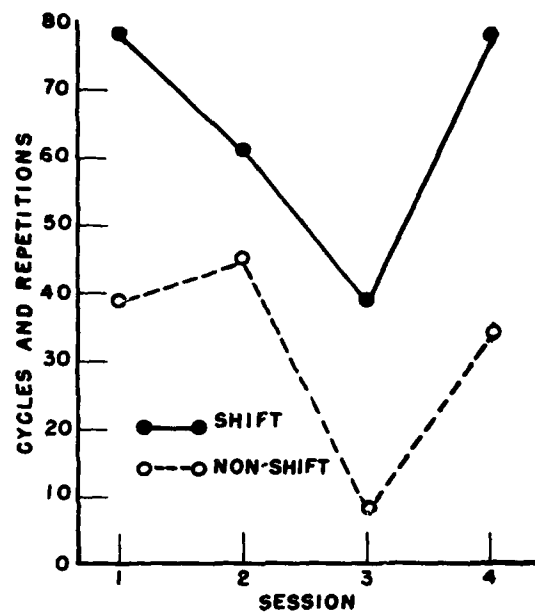


Figure 2 - Total number of cycles and repetitions produced by shift and non-shift subjects during each of four problem-solving sessions

363 incorrect) rule-line suggestions, and non-shift subjects made a total of 1399 (1229 correct and 170 incorrect) rule-line suggestions during the four problem sessions. Activity levels (total number of suggestions/time in min.) were computed for each subject for each session and added across the four sessions. The mean of the scores computed this way for non-shift subjects is 2.77; for shift subjects, 2.68. The corresponding value of t is not significant. The means of 2.77 and 2.68 are not directly interpretable, for they are based on sums of activity levels for four sessions. Actually subjects (shift and non-shift) made .55 rule-line suggestions per minute or approximately one every 2 min. of working time.

When times-to-solution for each of the 57 problems solved by non-shift subjects and 46 problems solved by shift subjects were averaged, the means were 9.16 and 11.80 min. respectively, although the CR of 1.83 is not statistically significant. The variance estimate for the non-shift groups is 39.99 and for the shift group is 65.76, yielding an F ($df = 45, 56$) of 1.64 ($p < .05$). We may thus conclude that the populations probably are different with respect to variability in times-to-solution. Fifty-five per cent of the non-shift solutions were obtained by the end of the first 6 min. on a problem; 35 per cent of the shift solutions were obtained by the end of 6 min. (or two 3-min. working periods) on a problem. Eighty-seven per cent of the non-shift solutions were obtained by the end of 15 min.; 70 per cent of shift solutions were obtained by the end of 15 min. No non-shift solutions were obtained in the last 3 min. of working time on problems; three shift solutions were obtained in that time interval.

Scores on the attitude questionnaires were obtained

as indicated in the previous section; 45 was the maximum score which a subject could make. The subjects in the non-shift group have a mean score of 39.9; subjects in the shift group have a mean score of 39.0. The difference is, of course, not significant. However, it is interesting to note that with respect to the only two items which specifically mention working on two problems during a session, responses of the two groups of subjects are distributed somewhat differently. The subjects in both groups worked on two problems per session, but under different conditions.

Item 11 reads as follows:

I felt that my work on one problem during a session

- _____ (a) definitely helped my performance on the other problem in the same session.
- _____ (b) did not markedly affect my performance on the other problem in the same session.
- _____ (c) definitely hurt my performance on the other problem in the same session.

The subjects' responses are distributed as shown in Table 3 (p. 15). $\chi^2 = 6$, $df = 2$. The probability of obtaining so large a value of χ^2 by chance is approximately .05.³

Item 14 on the questionnaire reads as follows:

If I had to work on problems like these again,

- _____ (a) I would want to work on two problems during one session as I did here.
- _____ (b) it would not make much difference to me whether I worked on two problems or one problem during one session.
- _____ (c) I would want to work on only one problem during one session.

³This takes into account Cochran's (3) recommendations concerning interpretation of χ^2 with small expected values. The smallest expectation in computing χ^2 from the table for Item 11 is 2.5.

The distribution of subjects' responses is shown in Table 3. With a χ^2 of 9.33 ($df = 2$), $p < .01$.

Table 3
Number of Subjects Marking Favorable (a), Neutral (b), and Unfavorable (c) Options to Two Questionnaire Items

| | Item 11 | |
|-----|---------|-----------|
| | Shift | Non-Shift |
| (a) | 5 | 7 |
| (b) | 6 | 9 |
| (c) | 5 | 0 |

$$\chi^2 = 6.00, df = 2$$

$$p < .05$$

| | Item 14 | |
|-----|---------|-----------|
| | Shift | Non-Shift |
| (a) | 4 | 8 |
| (b) | 3 | 7 |
| (c) | 9 | 1 |

$$\chi^2 = 9.33, df = 2$$

$$p < .01$$

DISCUSSION

Although subjects' over-all problem-solving performance (i.e., number of solutions, times) did not suffer seriously from imposed 3-min. shifts, interference effects were evidenced in terms of increases in cycles and repetitions. Such redundant activity is not necessary to the solution of the calculus problems. On the surface it would seem only to take time and make the problem-solving display more complex. In the face of failure to find a significant decrement in number of problems solved and an increment in solution times for shift subjects, however, we must conclude that some psychological advantages of redundancy outweighed its disadvantages. Perhaps redundant activity served to reinstate a tactical sequence interfered with by shift to another problem. For example, a subject working on Problem a

may have taken one or more steps consistent with his problem-solving hypothesis; he was then forced to shift to Problem b; on his return to Problem a after 3 min., some cue may have been necessary to reinstate the sequence of responses he had originally begun. A cycle or repetition may have fulfilled such a function.

Instances of lack of homogeneity of variance are of more than routine statistical interest in the present experiment. Significant differences in variability for shift and non-shift groups in terms of cycles, repetitions, errors, and times lead us to conclude that the shift procedure enhanced differences between individuals. With errors and times-to-solution, shift data offer evidence of more variability than non-shift data, even though group means do not differ significantly. It seems reasonable to hypothesize that a subject who could be expected to make more than the average number of errors working without interruption might make an even larger number of errors when he was forced to shift from one problem to another. Perhaps the poorer problem solver is especially sensitive to interference in the shift situation.

Total scores on the questionnaire used in the experiment are generally high and restricted to a narrow range for both groups of subjects. The experimenter informed the recruit subjects that their answers and scores would not be made available to their navy superiors. However, they either did not believe this or actually found the experimental sessions and attention a welcome reprieve from their rigid training schedules. They generally tended to mark "favorable" responses. Future work with similar subjects would certainly require a complete revision of the questionnaire, probably increasing the

specificity of items and the number of options, in order to increase the range of scores. Only then could meaningful between-group comparisons be made on the basis of over-all scores. This does not, however, discount the findings with respect to the only two items highly specific to working on two problems during a session (see p. 14). Responses to these items, interpreted in the light of the different problem solving conditions under which shift and non-shift subjects worked, indicate that shift subjects did not find their sessions so congenial as did non-shift subjects.

It is felt that the present experiment filled an initial need to go beyond showing that over-all performance is or is not affected when subjects are forced to shift from one problem-solving task to another. If only number of solutions and times had been recorded here, we would have been forced to conclude that there were no differences between shift and non-shift subjects. The data indicate, however, that in spite of the similarity with respect to over-all performance, shift subjects engaged in more redundant activity, were more variable with respect to certain measures, and probably felt less favorable toward the details of their problem-solving procedure than did subjects allowed to work on problems without interruption for relatively long periods of time. In the design of the experiment, however, many interesting problems were ignored. The data themselves suggested several more. For example, would shorter shift time intervals reduce over-all efficiency? What would happen if the number of problems were increased? An engineer at the Laboratory has suggested that perhaps, in relation to human "bandwidth," the "noise" introduced by the "50 per cent duty cycle" at the "rate" imposed here

does not even approach the value necessary to bring about a serious reduction in system "capacity." In the present experiment, a detailed analysis of problem-solving tactics (rule-line responses and sequences) was not carried out. Perhaps here is the real key to the effects of shift. Our conclusions point up the need for more knowledge about the role of redundant activity in problem solving and the factors tending to enhance individual differences in certain aspects of performance in a shift situation.

Navy - NRL, Bellevue, D. C.

REFERENCES

1. Bergstrom, J. A. Experiments upon physiological memory by means of the interference of associations. Amer. J. Psychol., 1892, 5, 356-369.
2. Bergstrom, J. A. The relation of the interference to the practice effect of an association. Amer. J. Psychol., 1893, 6, 433-442.
3. Cochran, W. G. The ² correction for continuity. Iowa State Coll. J. Sci., 1942, 16, 421-436.
4. Culler, A. J. Interference and adaptability. Arch. Psychol., 1912, 3, No. 24.
5. Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. J. exp. Psychol., 1948, 38, 404-411.
6. Harlow, H. F. The formation of learning sets. Psychol. Rev., 1949, 56, 51-65.
7. Jersild, A. T. Mental set and shift. Arch. Psychol., 1927, 14, No. 89.
8. Johnson, D. M. Problem solving and symbolic processes. In Stone, C. P. (Ed.), Annual review of psychology. Stanford, Calif.: Annual Reviews, Inc., 1950. Pp. 297-310.
9. Johnson, P. O. Statistical methods in research. New York: Prentice-Hall, 1949.
10. Moore, O. K., & Anderson, Scarvia B. Modern logic and tasks for experiments on problem solving behavior. J. Psychol., 1954, 38, 151-160.
11. Moore, O. K., & Anderson, Scarvia B. Search behavior in individual and group problem solving. Amer. sociol. Rev., 1954, 19, 702-714.
12. Youtz, R. P. The relation between number of confirmations of one hypothesis and the speed of accepting a new and incompatible hypothesis. Amer. Psychologist, 1948, 3, 248-249. (Abstract)